SYSTEMS FOR EXPRESSING TOXIC PROTEINS, VECTORS AND METHOD FOR PRODUCING TOXIC PROTEINS

DESCRIPTION

5 **TECHNICAL FIELD**

The present invention relates to systems for expressing toxic proteins, to expression vectors comprising one of these systems, to prokaryotic cells transformed with
10 these systems, and also to a method for synthesizing a toxic protein using these expression systems.

It enables, for example, the overproduction in a prokaryotic cell, for example *Escherichia coli*
15 *(E. coli)*, of toxic hydrophobic proteins or peptides, for example the overproduction of transmembrane domains of viral envelope proteins.

It finds many applications in particular in research
20 concerning the mechanisms of viral infections, and in the search for and development of novel active principles for combating viral infections.

In the description which follows, the references
25 between square brackets [ ] refer to the attached reference list.

State of the art

30 Determination of the three-dimensional (3D) structure is a decisive step in the structural and functional understanding of proteins.

Very great efforts and means have been, and are being,
35 used to achieve this aim, and have been amplified with

the accumulation of data provided by the genome sequencing programmes [1].

The two main techniques for establishing these protein
5   structures are X-ray diffraction, carried out using crystallized proteins, and nuclear magnetic resonance (NMR) carried out using proteins in solution. NMR, which is very suitable for studying proteins with a molecular mass of less than 20 kDa, requires however,
10   like X-ray diffraction, the production of large amounts of material. It also means, in most cases, that material enriched in ^{15}N and/or ^{13}C must be prepared.

In this context, the bacterium is a means of production
15   that is widely used by the scientific community [2]. The overexpression of proteins in bacteria does not, however, occur without problems. In fact, it gives rise to three situations:
-    The first case, which is ideal, is that where the
20   protein is overproduced in a form that is correctly spatially folded during its synthesis in vivo. This is not a rare situation, but neither is it frequent. It concerns essentially soluble proteins that are small, i.e. approximately 20 to 50 kDa.
25

-    The second case, the most common, is that where the protein is overproduced and aggregated in the form of inclusion bodies. This concerns polytopic and/or large proteins. In this case, the kinetics of folding
30   of the protein are clearly slower than its rate of biosynthesis. This promotes exposure of the hydrophobic regions of the protein, that are normally buried in the core thereof, to the aqueous solvent and generates non-specific interactions that result in the formation of
35   insoluble aggregates. According to the degree of

disorder of this folding, the inclusion bodies can be solubilized/unfolded under non-native conditions, with urea or guanidine. The solubilized protein is then subjected to various treatments, such as dialysis or dilution, so as to promote, successfully in certain cases, a native 3D folding.

-    The third case is that where the expression engenders a varying degree of toxicity. This goes from an absence of expression product if the bacterium manages to adapt itself, to death of the bacterium if the product is too toxic. It is a case which occurs quite frequently and most commonly with membrane proteins or membrane protein domains, for instance those of the envelope proteins of the hepatitis C virus [5] or of the human immunodeficiency virus [6].

The problem of toxicity relates essentially to the expression of membrane proteins, i.e. proteins having a hydrophobic domain. Now, these proteins are of growing interest. Firstly, they are relatively numerous since the establishment of the various genomes confirms that they represent approximately 30% of the proteins potentially encoded by these genomes [7]. Secondly, they constitute 70% of the therapeutic targets and their alteration is the cause of many genetic diseases [8].

It is therefore essential to develop methods that facilitate or allow the expression of such proteins or of their membrane portion.

Efforts have been made in this respect with, for example, the development of bacterial strains that either show better tolerance to the expression of

membrane proteins [9, 10], or have a stricter regulation of the mechanism in the expression, as in the case of the E. coli strain BL21(DE3)pLysS developed by Stratagene. However, these improvements do not make
5   it possible to eliminate the toxicity phenomenon in all cases, in particular in the expression of hydrophobic peptides corresponding to membrane anchors.

The treatment of hepatitis C currently represents one
10   of the major high-stakes areas of medicine. Hepatitis C is caused by the hepatitis C virus (HCV) of the family of flaviviridae and which specifically infects hepatic cells [11]. This virus consists of a positive RNA of approximately 9500 bases which encodes a polyprotein of
15   3033 residues [13], symbolized in the attached Figure 1 by the rectangle 1A. This polyprotein is cleaved, after expression, by endogenous and exogenous proteases, so as to give rise to 10 different proteins. Two of them, called E1 and E2, are glycosylated and form the
20   envelope of the virus. They each have membrane domains called TM, in particular TME1 for the E1 protein and TME2 for the E2 protein. The cleavage positions that generate them are indicated in Figure 1 by arrows with, mentioned below, a number which corresponds to the
25   position in the polyprotein of the first amino acid of sequence resulting from the cleavage. The E1 and E2 proteins are symbolized by a rectangle. The white portion of each rectangle corresponds to the ectodomain (ed) and the shaded domain to the transmembrane region
30   (TM). The primary sequence of the TMs is indicated at the bottom of the figure in one—letter—code, with numbers corresponding to the position of the amino acids in the polyprotein located at the ends of these domains. The stars indicate the hydrophobic amino
35   acids. These membrane domains or membrane regions of

the virus have particular association properties that condition the structuring of the viral envelope [12]. In this respect, they constitute potential therapeutic targets. An understanding of the mechanism of association of the virus requires studies of the 3D structure of these domains, in particular by means of the abovementioned techniques, which involves producing these peptides in abundant amounts, and also preferably via the biosynthetic pathway in order to allow ^{15}N and/or ^{13}C isotope labelling.

The various E1 expression trials of the prior art, in particular in *E. coli* [14] [5] or in *sf9* insect cells infected with baculoviruses [15], have not made it possible to overproduce this E1 protein, in particular due to the toxicity induced by its expression, including in the "resistant" *E. coli* BL21(DE3)pLysS strains described above. There has been no E2 protein overexpression trial in bacteria. These toxicity problems are essentially due to the C-terminal region of the two proteins, that is rich in hydrophobic amino acids which form transmembrane domains that provide the anchoring to the membrane of the endoplasmic reticulum.

There is therefore a real need for a system for expressing toxic proteins which does not have the drawbacks, and limitations, deficiencies and disadvantages of the techniques of the prior art.

In addition, there is a real need for an expression vector comprising such a system for expressing toxic proteins, making it possible to carry out a method for producing toxic proteins which does not have the drawbacks, limitations, deficiencies and disadvantages of the techniques of the prior art.

DISCLOSURE OF THE INVENTION

The aim of the present invention is precisely to provide a system for expressing a toxic protein, which satisfies, inter alia, the needs indicated above.

This aim, and others, are achieved, in accordance with the invention, by means of an expression system characterized in that it comprises successively, in the 5'-3' direction, a nucleotide sequence encoding the dipeptide Asp-Pro, referred to below as *dp* sequence, and a nucleotide sequence (*pt*) encoding a toxic protein (Pt). This system will be identified below by: *dp-pt*.

According to a particularly preferred embodiment of the present invention, the expression system also comprises, upstream of the *dp* sequence, a nucleotide sequence (*ps*) encoding a soluble protein (Ps). This soluble protein may be, for example, glutathione S-transferase (GST) or thioredoxin (TrX) or another equivalent soluble protein. This expression system according to the invention will be identified below by: *ps-dp-pt*.

The *dp-pt* expression system of the present invention, which comprises a sequence encoding Asp-Pro (DP in one-letter code) placed upstream of the nucleotide sequence of the toxic protein, makes it possible, entirely unexpectedly, to suppress the toxic effect of the protein for the host cell. In addition, the inventors have noted that, entirely surprisingly, the suppression of toxicity of the protein in the host is even more effective with the *ps-dp-pt* expression system when the toxic peptide is produced as a C-terminal fusion with a

soluble protein, for example glutathione S-transferase or thioredoxin, with the sequence Asp-Pro inserted between the soluble protein and the toxic peptide.

5   The *dp-pt* or *ps-dp-pt* expression system of the present invention makes it possible to overproduce toxic proteins in host cells, in particular hydrophobic proteins, especially peptides which correspond to, or which comprise, hydrophobic domains of membrane-
10   anchored proteins which may involve, for example, a membrane protein or a domain of a membrane protein. It may involve, for example, a protein of a virus, for example of a hepatitis C virus, of an AIDS virus, or of any other virus that is pathogenic for humans and, in
15   general, for mammals.

For example, the *dp-pt* or *ps-dp-pt* system of the invention makes it possible to overproduce, in a host such as *E. coli*, the transmembrane domains of the E1
20   and E2 proteins of the hepatitis C virus, called TME1 and TME2, corresponding respectively to the sequences:

TME1:   347-MIAGAHWGVLAGIAYFSMVGNWAKVLVVLLLFAGVDA-383
sequence ID No. 1

25

TME2:   717-MEYVVLLFLLLADARVCSCLWMMLLISQAEA-746
Sequence ID No. 2

whereas this was not possible with the techniques of
30   the prior art.

The nucleotide sequences that can be used for constituting the *dp-pt* system of the invention encoding the TME1 ($dp-pt_{(TME1)}$) or TME2 ($dp-pt_{(TME2)}$) proteins can
35   be any of the possible sequences encoding respectively

the DP-TME1 and DP-TME2 fusion proteins. The sequences encoding the TME1 and TME2 proteins may advantageously be, for example, sequence ID No. 3 and sequence ID No. 4, respectively, of the attached sequence
5   listing. To obtain the *dp-pt* system, the *dp* sequence encoding the dipeptide Asp-Pro (DP) is added to these sequences.

The nucleotide sequences that can be used for
10   constituting the *ps-dp-pt* system of the invention encoding the TME1 (*ps-dp-pt*$_{(TME1)}$) or TME2 (*ps-dp-pt*$_{(TME2)}$) proteins may be any of the possible sequences encoding the Ps-DP-TME1 and Ps-DP-TME2 fusion proteins, respectively. They may advantageously be, for example,
15   the sequences ID No. 34, ID No. 35 and ID No. 36 of the attached sequence listing for TME1, making it possible to obtain a Ps-DP-TME1 chimeric protein. They may advantageously be, for example, the sequences ID No. 37, ID No. 38 and ID No. 39 of the attached
20   sequence listing for TME2, making it possible to obtain a Ps-DP-TME2 chimeric protein.

In fact, the abovementioned nucleotide sequences have optimized codons for the expression of TME1 and TME2 in
25   a bacterium, for example in *E. coli*.

A large number of HCV RNA sequences producing an infectious phenotype exist: these sequences can also be used in the present invention.
30

The sequence encoding the dipeptide Asp-Pro may be, for example: gacccg, or any other sequence encoding this dipeptide.

35   The sequence encoding GST may be, for example, that

present in the pGEXKT plasmids, the sequence of which corresponds to sequence ID No. 29 of the attached sequence listing, or any equivalent sequence, i.e. encoding this soluble protein. The sequence encoding

5   TrX may be, for example, that present in the pET32a+ expression plasmid, the sequence of which corresponds to sequence ID No. 30 of the attached sequence listing, or any equivalent sequence, i.e. encoding this soluble protein.

10

For the production of the toxic protein, the *dp-pt* or *ps-dp-pt* expression system of the invention is placed inside a host cell, for example by cloning in an appropriate plasmid, by means of the usual techniques

15 for transforming a host in genetic recombination techniques.

The plasmid into which the expression system of the present invention may be cloned so as to form this

20 vector will be chosen in particular according to the host cell. It may be, for example, the pT7-7 plasmid (sequence ID No. 33 of the attached sequence listing), a plasmid of the pGEX series (for example of sequence ID No. 31 of the attached sequence listing), sold for

25 example by the company Pharmacia, or a plasmid of the pET32 series (for example of sequence ID No: 32 of the attached sequence listing), sold for example by the company Novagen.

30 The plasmids of the pGEX series and of the pET32 series will advantageously be used for implementing the present invention. In fact, they already comprise a *ps* sequence encoding a soluble protein (Ps), respectively glutathione S-transferase and thioredoxin. Thus,

35 advantageously, the *dp-pt* system will be cloned into

these plasmids downstream of this *ps* sequence encoding
the soluble protein.

The present invention therefore also relates to an
expression vector comprising a *dp-pt* or *ps-dp-pt*
expression system according to the invention; in
particular, a vector comprising a *dp-pt* expression
system according to the invention and the
oligonucleotide sequence of the pT7-7 plasmid, or a
vector comprising a *ps-dp-pt* expression system
according to the invention and the oligonucleotide
sequence of a pGEX plasmid or of a pET32 plasmid.

For example, the expression vectors of the present
invention that are suitable for a bacterial host such
as *E. coli* and that allow overexpression of the
abovementioned TME1 membrane protein may advantageously
have an oligonucleotide sequence chosen from the
sequences ID No. 40 (with pGEXKT), ID No. 42 (with
pET32a+) and ID No. 44 (with PT7-7) of the attached
sequence listing.

For example, the expression vectors of the present
invention that are suitable for a bacterial host such
as *E. coli* and that allow overexpression of the
abovementioned TME2 membrane protein may advantageously
have an oligonucleotide sequence chosen from the
sequences ID No. 41 (with pGEXKT), ID No. 43 (with
pET32a+) and ID No. 45 (with pT7-7) of the attached
sequence listing.

In fact, the abovementioned expression vectors have
codons that are optimized for the expression of the
chimeric proteins of the present invention, including
TME1 and TME2, in a bacterium, for example in *E. coli*.

The present invention also relates to a prokaryotic
cell transformed with an expression vector according to
the invention. This prokaryotic cell transformed with
the expression vector of the present invention should
preferably allow overexpression of the toxic protein
for which the vector codes. Thus, any host cell capable
of expressing the expression vector of the present
invention can be used, for example *E. coli*,
advantageously the *E. coli* strain BL21(DE3)pLysS.

The present invention also relates to a method for
producing a toxic protein by genetic recombination,
comprising the following steps:

—   transforming a host cell with an expression vector
according to the invention,
—   culturing the transformed host cell under culture
conditions such that it produces a fusion protein
comprising the dipeptide Asp-Pro followed by the
peptide sequence of the toxic protein from said
expression vector, and
—   isolating said fusion protein, and
—   cleaving said fusion protein so as to recover the
toxic protein.

The steps for transforming, culturing and isolating the
chimeric protein produced can be carried out by means
of the usual techniques of genetic recombination, for
example by means of techniques such as those that are
described in document [25].

The step consisting in isolating the fusion protein can
be carried out by means of the usual techniques known
to those skilled in the art for isolating a protein

from a cell extract.

The fusion protein produced by means of the method of the invention has a "soluble protein-Asp-Pro-toxic protein" sequence. In the present description, the dipeptide Asp-Pro is also called DP according to the one-letter amino acid code.

For example, when the toxic protein is TME1, the fusion protein may have the sequence ID No. 46 of the attached sequence listing, which corresponds to the GST-DP-TME1 fusion protein; the sequence ID No. 48 of the attached sequence listing, which corresponds to the TrX-DP-TME1 fusion protein; or the sequence ID No. 50 of the attached sequence listing, which corresponds to the M-DP-TME1 fusion protein of the attached sequence listing.

For example, when the toxic protein is TME2, the fusion protein may have the sequence ID No. 47 of the attached sequence listing, which corresponds to the GST-DP-TME2 fusion protein; the sequence ID No. 49 of the attached sequence listing, which corresponds to the TrX-DP-TME2 fusion protein; or the sequence ID No. 51 of the attached sequence listing, which corresponds to the M-DP-TME2 fusion protein of the attached sequence listing.

The step consisting of cleavage of this fusion protein can advantageously be carried out by means of formic acid, which cleaves the fusion protein at the dipeptide Asp-Pro. It may be carried out, moreover, by means of any appropriate technique known to those skilled in the art for recovering a protein from a sample using a fusion protein.

The inventors are the first to have found a system that is really effective for producing and even overproducing, in particular in the *Escherichia coli (E. coli)* bacterium, hydrophobic peptides corresponding to the membrane domains of the E1 and E2 proteins of the hepatitis C virus envelope, the expression of which is lethal for the microorganism.

The field of application of the present invention concerns mainly the production of hydrophobic peptides on a large scale, in particular for fundamental and industrial research. In addition, the production of the chimeric protein consisting of the soluble protein, of the dipeptide Asp-Pro and of the hyrophobic peptide can be used for a functional purpose, in particular for obtaining information on the degree of oligomerization of the membrane domain or else on its heteropolymerization capacity.

The fusion proteins, or chimeric proteins, are produced via their coding DNA present, for example, in commercial plasmids and following which is introduced, in phase, the DNA encoding the Asp-Pro sequence followed by that encoding the toxic peptide. This application can be commercialized in the form of bacterial expression plasmids which will include the sequence of the Asp-Pro site, downstream of that of the soluble proteins already present. The corresponding plasmid will be described, for example, as a tool that facilitates the production, via the biological pathway, of toxic membrane peptides or proteins.

Thus, the present invention is applicable to any system for overexpressing recombinant proteins, with or

without fusion to a soluble protein such as, for example, GST or thioredoxin, including a non-natural Asp-Pro sequence inserted upstream of a sequence encoding a toxic domain of the protein, for example a
5   membrane domain of a protein.

Other characteristics and advantages of the present invention will become further apparent to those skilled in the art on reading the following examples given by
10   way of non-limiting illustration, with reference to the sequence listing and to the figures that are attached.

Brief description of the attached sequence listing

15   -   Sequences ID Nos. 1 and 2: peptide sequences of
          TME1 and of TME2, respectively.
     -    Sequences ID Nos. 3 and 4: sequences encoding the
          TME1 peptide and the TME2 peptide, respectively.
     -    Sequences ID Nos. 5 and 6: respectively,
20   oligonucleotide (+) for insertion into pT7-7
          (OL13(+)) and oligonucleotide (-) for insertion
          into pT7-7 (OL14(-)).
     -    Sequences ID Nos. 7 and 8: respectively, coding
          sense DNA of TME1 + *cla* I site in the 3' position
25   and anticoding sense DNA of TME1 + *cla* I site in
          the 5' position (sequence complementary to the
          sequence ID No. 7).
     -    Sequences ID Nos. 9 and 10: respectively, coding
          sense oligonucleotide (OL11(+)) and anticoding
30   sense oligonucleotide (OL12(-)) for the synthesis
          of TME1.
     -    Sequence ID No. 11: oligonucleotide (+) for
          insertion into pGEXKT without dp site (OL15(+)).
     -    Sequence ID No. 12: oligonucleotide (+) for
35   insertion into pGEXKT with dp site (OL17(+)).

- Sequence ID No. 13: oligonucleotide (-) for insertion into pGEXKT (OL16(-)).
- Sequence ID No. 14: oligonucleotide (+) for insertion into pET32a (OL18(+)) (hybridizes to the segment 915-932 of pGEXKT).
- Sequences ID Nos. 15 and 16: respectively, oligonucleotides (+) (OL19(+)) and (-) (OL20(-)) for insertion into pT7-7 of the DNA encoding MDP-TME1.
- Sequences ID Nos. 17 and 18: respectively, oligonucleotide (+) for insertion into pT7-7 (OL23(+)) and oligonucleotide (-) for insertion into pT7-7 (OL24(-)).
- Sequences ID Nos. 19 and 20: respectively, coding sense DNA for TME2 + *Nde* I site in the 5' position and *Hind* III site in the 3' position; and anticoding sense DNA of TME2 + *Nde* I site in the 3' position and *Hind* III site in the 5' position (sequence complementary to ID No. 17).
- Sequences ID Nos. 21 and 22: respectively, coding sense oligonucleotide (OL21(+)) and anticoding sense oligonucleotide (OL22(-)) for the synthesis of TME2.
- Sequence ID No. 23: oligonucleotide (+) for insertion into pGEXKT without dp site (OL25(+)).
- Sequences ID Nos. 24 and 25: respectively, oligonucleotides (+) (OL27(+)) and (-) (OL26(-)) for insertion into pGEXKT with dp site.
- Sequences ID Nos. 26 and 27: respectively, oligonucleotides (+) (OL28(+)) and (-) (OL29(-)) for insertion into pT7-7 of the DNA encoding MDP-TME2.
- Sequence ID No. 28: end of the sequence of the GST soluble protein followed by the thrombin site encoded in the pGEXKT plasmid.

         –   Sequence ID No. 29: DNA encoding the GST protein in the pGEXKT plasmid.

         –   Sequence ID No. 30: DNA encoding thioredoxin (TrX) in the pET32a+ plasmid.

5   –   Sequences ID Nos. 31, 32 and 33: respectively, pGEXKT, pET32a+ and pT7-7 expression plasmids.

         –   Sequences ID Nos. 34, 35 and 36: respectively, expression systems according to the invention encoding the GST-DP-TME1, TrX-DP-TME1 and

10   M-DP-TME1 fusion proteins.

         –   Sequences ID Nos. 37, 38 and 39: respectively, expression systems according to the invention encoding the GST-DP-TME2, TrX-DP-TME2 and M-DP-TME2 fusion proteins.

15   –   Sequences ID Nos. 40 and 41: respectively, pGEXKT-dp-pt_{TME1} and pGEXKT-dp-pt_{TME2} expression vectors according to the invention encoding the GST-DP-TME1 and GST-DP-TME2 fusion proteins.

         –   Sequences ID No. 42 and 43: respectively,

20   pET32a-dp-pt_{TME1} and pET32a-dp-pt_{TME2} expression vectors according to the invention encoding the TrX-DP-TME1 and TrX-DP-TME2 fusion proteins (code via the complementary strand).

         –   Sequences ID Nos. 44 and 45: respectively,

25   pT7-7-dp-pt_{TME1} and pT7-7-dp-pt_{TME2} expression vectors according to the invention encoding the MDP-TME1 and M-DP-TME2 fusion proteins.

         –   Sequences ID Nos. 46 and 47: respectively, GST-DP-TME1 and GST-DP-TME2 fusion proteins

30   according to the invention obtained from the pGEXKT-dp-pt_{TME1} and pGEXKT-dp-pt_{TME2} plasmids.

         –   Sequences ID Nos. 48 and 49: respectively, TrX-DP-TME1 and TrX-DP-TME2 fusion proteins according to the invention obtained from the

35   pET32a-dp-pt_{TME1} and pET32a-dp-pt_{TME2} plasmids.

- Sequences ID Nos. 50 and 51: respectively, M-DP-TME1 and M-DP-TME2 fusion proteins according to the invention obtained from the pT7-7-*dp-pt*TME1 and pT7-7-*dp-pt*TME2 plasmids.

5 - Sequences ID Nos. 52 and 53: respectively, GST and TrX proteins encoded by the pGEXKT and pET32a+ vector.

Brief description of the figures

10

- Figure 1: diagrammatic representation of a portion of the HCV polyprotein and peptide sequence of the C-terminal membrane domains of the E1 and E2 envelope proteins. The peptide sequences represented correspond
15 to the infectious type #D00831 and #M67463 for TME1 and TME2, respectively, obtained from the public sequence library of the European Molecular Biology Laboratory (EMBL).

20 - Figure 2: creation of the DNA encoding the C-terminal membrane domain of the HCV E1 envelope protein and additional sequences in the 5' and 3' positions for cloning in various plasmids. The sequences represented in this figure are reported in
25 the attached sequence listing.

- Figure 3: creation of the DNA encoding the C-terminal membrane domain of the HCV E2 envelope protein and additional sequences in the 5' and 3'
30 positions for cloning in various plasmids. The sequences represented in this figure are reported in the attached sequence listing.

- Figure 4, panels A to F: toxicity of the membrane
35 domains expressed in the bacterium and suppression of

this toxicity by insertion of a dp site. Panels A, C and E are graphic representations of optical density (OD) measurements at 600 nm as a function of time (t) in hours of production of various proteins in a bacterium using or not using the expression system of the present invention. Panels B, D and F are representations of the gels of migration of the proteins of panels A, C and E, respectively.

-     Figures 5A and B: overexpression of the thioredoxin-Asp-Pro-Pt chimeric proteins (Pt = membrane domains of the proteins) in the bacterium. Figure 5A is a graphic representation of the optical density (OD) measurements at 600 nm as a function of the time in hours of production of various proteins in a bacterium using or not using the expression system of the present invention: Figure 5B is a representation of a gel of migration of the proteins of Figure 5A.

-     Figure 6: expression and purification of the GST-TME2 fusion (or chimeric) protein, and comparison with GST alone. This figure represents, at the top, the peptide sequences of GST and GST-TME2, and, at the bottom, the gels obtained by electrophoresis, showing that, unlike GST alone, GST-TME2 is insoluble. The latter is produced in the form of inclusion bodies that cannot fold correctly.

-     Figures 7A and 7B: graphic representations of comparative experimental results showing the effect of the DP dipeptide (*dp-pt* oligonucleotide sequence in accordance with the present invention) and of the DP dipeptide and the soluble protein (*ps-dp-pt* oligonucleotide sequence in accordance with the present invention) on the synthesis of the TME1 and TME2 toxic proteins in accordance with the present invention.

EXAMPLES

In these examples, the oligonucleotides used were ordered from Laboratoires EUROBIO (http://www.eurobio.fr/); the plasmids were prepared with the QIAprep kit (brand name) from Qiagen (http://www.qiagen.com/); the DNA sequences were sequenced with the ABI PRISM (registered trade mark) BigDye (brand name) Terminator cycle kit from Applied Biosystems (http://home.appliedbiosystems.com/); the *E. coli* strains BL21(DE3) and BL21(DE3)pLysS were obtained from Stratagene (http://www.stratagene.com/); the C41 and C43 (BL21(DE3)) strains were provided by Dr. Bruno Miroux (CNRS-CEREMOD, Centre for Research on molecular endocrinology and development; the DNA restriction and modification enzymes were obtained from New England Biolabs (http://www.neb.com/neb/); the protein electrophoreses were carried out with a miniprotean 3 (brand name) from Bio-Rad Laboratories (http://www.bio-rad.com); the plasmid pCR (registered trade mark) T7 topo TA was obtained from Invitrogen (http://www.invitrogen.com/); the pET32a+ plasmid was obtained from Novagen (http://www.novagen.com); the pT7-7 and pGP1-2 plasmids and the K38 strain [22] were requested from Prof. Tabor (Department of Biological Chemistry, Harvard Medical School); the pGEX-KT plasmid was requested from Prof. Dixon (Department of Biological Chemistry, University of Michigan Medical School); the other products were obtained from Sigma (http://sigma.aldrich.com).

In the following examples, the production of the TME1 and TME2 peptides was firstly carried out without the expression system of the present invention, and then as a fusion with a soluble protein and, finally, as a

fusion with GST with insertion of the Asp-Pro ("DP" in one-letter coding) site between the soluble protein and TME1 or TME2.

5   The abbreviation "SEQ ID No." is used for "sequence ID No." and refers to the attached sequence listing.

Example 1: Synthesis of the expression system

1.1) <u>CONSTRUCTION OF THE pT7-7-pt_{TME1} and pT7-7-pt_{TME2}</u>
10   <u>EXPRESSION VECTORS</u>

The DNA encoding the two domains was synthesized *de novo* using the appropriate oligonucleotides. The codons were chosen according to their greatest
15   frequency of use in the bacterium, as was quantified by Sharp et al. [17]. The constructs are described in the attached Figure 2 for TME1 and in the attached Figure 3 for TME 2.

20   Each synthetic DNA was generated using a set of two long and overlapping oligonucleotides, OL11 (SEQ ID No. 9) and OL12 (SEQ ID No. 10) for TME1, and OL21 (SEQ ID No. 19) and OL22 (SEQ ID No. 20) for TME2, which were amplified after hybridization with two external
25   oligonucleotides chosen according to the cloning in a given plasmid. Thus, the clonings in pT7-7 were carried out using the set of external oligonucleotides OL13 (SEQ ID No. 5) and OL14 (SEQ ID No. 6) for TME1, and OL23 (SEQ ID NO. 15) and OL24 (SEQ ID NO. 16) for TME2.
30

Each synthetic DNA was generated using a set of four oligonucleotides: two long and overlapping and two short and external. The DNAs were amplified by the polymerase chain reaction method, referred to as "PCR"
35   [18], and then cloned into a bacterial plasmid pCR

(brand name) T7 topo TA. The synthesized DNAs were sequenced and then subcloned into the pT7-7 bacterial expression vector [19] using the *Nde* I restriction site in the 5' position and the *Cla* I or *Hind* III
5   restriction site in the 3' position.

In Figure 2:

A: TME1 peptide sequence of subtype #D00831. The
10   numbering corresponds to the position of the sequence in the polyprotein as described in Figure 1.
B: DNA sequence encoding the membrane domain with optimized codons for expression in the bacterium.
C and **D:** Strategy for DNA amplification without
15   matrix. The coding sense and the anticoding sense of the oligonucleotides are indicated, respectively, by the signs (+) and (-). The long oligonucleotides overlap by about twenty bases so as to create the primer and then the matrix. The short oligonucleotides
20   make it possible to amplify the matrix by PCR, integrating the desired restriction sites according to the plasmids used. The insertion into pT7-7 was carried out with the pair of oligonucleotides OL13 (SEQ ID No. 5) and OL14 (SEQ ID No. 6), via a subcloning in
25   pCRT7 topo, integrating the *Nde* I and *Hind* III sites. The insertion into pGEXKT was carried out according to the same method, with the pair of oligonucleotides OL15 (SEQ ID No. 11) and OL16 (SEQ ID No. 13), integrating the *BamH* I and *EcoR* I sites. The insertion of the dp
30   site (gacccg) and the cloning in pGEXKT were carried out with the pair of oligonucleotides OL17 (SEQ ID No. 12) and OL16 (SEQ ID No. 13). The construct in pGEXKT was transferred into pET32a, which encodes thioredoxin, with the pair of oligonucleotides OL18
35   (SEQ ID No. 14) and OL16 (SEQ ID No. 13). The

oligonucleotide OL18 (SEQ ID No. 14) hybridizes in the terminal region of the DNA encoding GST in pGEXKT. The amplified sequence integrates the end of GST (SDLSGGGGG) followed by the thrombin site (LVPRGS) (SEQ

5   ID No. 28), by the DP site and by the membrane passage. After cloning, the DNA inserted into pET32a makes it possible to express the thioredoxin-SDLSGGGGGLVPRGS-DP-TME1 chimera (SEQ ID NO. 48).

10   In Figure 3:

The legend is identical to Figure 2, but the peptide sequence is that of subtype #M67463. The insertion into pT7-7 was carried out with the pair of oligonucleotides

15   OL23 and OL24 (SEQ ID NO. 17 and SEQ ID No. 18, respectively), via a subcloning in pCRT7 topo, integrating the *Nde* I and *Hin*d III sites.

The insertion into pGEXKT was carried out according to

20   the same method, with the pair of oligonucleotides OL25 and OL26 (SEQ ID No. 23 and SEQ ID No. 25, respectively), integrating the *Bam*H I and *Eco*R I sites. Insertion of the dp site (gacccg) and the cloning in pGEXKT were carried out with the pair of

25   oligonucleotides OL27 and OL26 (SEQ ID No. 24 and SEQ ID No. 25, respectively). Insertion into pET32a was carried out as described in Figure 2, using the pair of oligonucleotides OL18 and OL26 (SEQ ID No. 14 and SEQ ID No. 25, respectively).

30

1.2) CONSTRUCTION OF THE pGEXKT-*pt*$_{TME1}$, pGEXKT-*pt*$_{TME2}$, pGEXKT-*dp-pt*$_{TME1}$ AND pGEXKT-*dp-pt*$_{TME2}$ EXPRESSION VECTORS

The pGEXKT-*pt*$_{TME1}$ and pGEXKT-*pt*$_{TME2}$ expression vectors

35   were constructed by PCR as described in the attached

Figures 2 and 3. The matrix DNA used to amplify the DNAs encoding TME1 or TME2 is that cloned into the pT7-7 plasmids. The cloning of TME1 into the pGEXKT plasmid [20, 21] was carried out using the sets of
5   oligonucleotides OL15 (SEQ ID No. 11) and OL16 (SEQ ID No. 13) allowing insertion of the *BamH* I restriction site in the 5' position and the *EcoR* I restriction site in the 3' position. The cloning of TME2 into the same vector was carried out using the sets of
10   oligonucleotides OL25 (SEQ ID No. 21) and OL26 (SEQ ID No. 23).

As indicated in Figure 2, the insertion of the *dp* site at the N-terminal position of TME1 was carried out by
15   replacing the 5' oligonucleotide OL15 (SEQ ID No. 11) with the oligonucleotide OL17 (SEQ ID No. 12). The insertion of the *dp* site at the N-terminal position of TME2 was carried out by replacing the 5' oligonucleotide OL25 (SEQ ID No. 21) with the
20   oligonucleotide OL27 (SEQ ID No. 22), as shown in Figure 3.

1.3) CONSTRUCTION OF THE pET32a-dp-TME1 AND pET32a-dp-TME2 EXPRESSION VECTORS

25

The pET32a-dp-TME1 and pET32a-dp-TME2 expression vectors were constructed by PCR as described in the attached Figures 2 and 3, using the set of oligonucleotides indicated. The upstream
30   oligonucleotide integrates an *EcoR* V site and hybridizes with the terminal region of the gene encoding GST. It makes it possible to integrate the 5-glycine tail and the thrombin-cleavage site present in the plasmid. The downstream oligonucleotide is the
35   same as that used for the cloning in pGEXKT.

The insertion into the pET32a plasmid is carried out via the *Ms*C I/*Eco*R V sites in the 5′ position and the *Eco*R I site in the 3′ position. It makes it possible to insert, in phase at the end of the thioredoxin sequence, the 5-glycine tail, the thrombin-cleavage site, the DP site and the membrane passage. The pET32a plasmid of origin, which serves as a control, encodes thioredoxin followed by a sequence integrating various elements that have not been deleted and that contribute, to a large degree, to the mass of the chimeric protein produced.

The matrix DNA used to amplify the DNAs encoding TME1 or TME2 is that cloned into the pGEXKT-*dp-pt*$_{TME1}$ or pGEXKT-*dp-pt*$_{TME2}$ plasmids. For TME1, the cloning into pET32a+ was carried out using the sets of oligonucleotides OL18 (SEQ ID No. 14) and OL16 (SEQ ID No. 13). The cloning of TME2 into the same vector was carried out using the sets of oligonucletides OL18 (SEQ ID No. 14) and OL26 (SEQ ID No. 23), as indicated in Figure 3.

Example 2: Expression of sequences encoding the TME1 and TME2 proteins alone

The expression of the sequences encoding the TME1 and TME2 domains alone was tested by thermal or chemical induction and using various bacterial strains as described below.

2.1) THERMAL INDUCTION SYSTEM

The system developed by Tabor [22] makes it possible to express a protein by thermal induction using two vectors in the same bacterium, pT7-7 and pGP1-2.

The pT7-7 plasmid contains the DNA to be expressed, placed under the control of a ϕ10 promoter recognized by the T7 phage RNA polymerase. The pGP1-2 plasmid contains the gene encoding the T7 phage polymerase, placed under the control of a λp_L promoter. This promoter is repressed by a thermosensitive repressor, cI857, that is itself also present in pGP1-2. At 30°C, cI857 is normally expressed and represses the λp_L promoter, which blocks the expression of the polymerase and therefore also that of the protein of interest.

The induction is triggered by switching the culture from 37 to 42°C for 15-30 min, and then the expression continues at 37°C. This system is therefore particularly suitable when it is necessary to strictly control the expression of a given protein, in particular if said protein is toxic for the bacterium.

2.2 CHEMICAL INDUCTION SYSTEM

The same pT7-7 plasmid containing the DNA to be expressed is this time introduced into *E. coli* bacteria of the type BL21(DE3) (B f⁻ dcm omtP hsdS$(r_B^- m_B^-)$ gal λ (DE3)) and BL21(DE3)pLysS (B F⁻ dcm ompT hsdS$(r_B^- m_B^-)$ gal λ (DE3) [pLysS Cam^r]). These bacteria have been modified so as to contain in the genome a copy of the gene encoding the T7 phage RNA polymerase, placed under the control of a lacUV5 promoter that can be induced with isopropyl-1-thio-β-D-galactoside (IPTG). In this case, the bacteria are cultured at their optimum temperature of 37°C or less if necessary. The expression is induced by adding IPTG to the culture. The BL21(DE3)pLysS strain is particularly suitable for proteins whose base line expression is toxic for the

host bacterium. In fact, the presence of the pLysS plasmid allows continuous expression, at a low level, of T7 phage lysozyme. This inhibits the T7 phage polymerase, the weak expression of which in the absence of induction could allow the base line expression of toxic protein.

The inventors also tested the expression of the membrane domains alone in strains called C41 and C43 [10], which were selected so as to withstand the expression of toxic membrane proteins. These strains are derived from the BL21(DE3) strain and are used in the same way as the latter.

2.3) EXPRESSION TESTS

According to the system tested, the corresponding plasmids were introduced by transformation into the various strains of *E. coli*: K38 (HfrC λ) for the Tabor thermal induction system or the various BL21 strains for the chemical induction. Table 1 below summarizes the tests performed.

Table 1

Induction	Strain	Plasmid
Thermal	K38	pT7-7+pGP1-2
Chemical	BL21(DE3)	pT7-7
Chemical	BL21(DE3)pLysS	pT7-7
Chemical	C41(BL21(DE3))	pT7-7
Chemical	C43(BL21(DE3))	pT7-7

In each case, about ten transformants were placed in culture in order to test the expression. Briefly, the bacteria were cultured in 5 ml of LB (10 g tryptone, 5 g yeast extract, 5 g NaCl, qs 1 litre H_2O),

supplemented with 50 µg/ml of ampicillin (necessary in order to maintain pT7-7 in the bacterium) and 60 µg/ml of kanamycin (necessary in order to maintain pGP1-2 in the bacterium), and then cultured until saturation,

5   either at 30°C for K38 or at 37°C for BL21(DE3). The cultures were then diluted to 1/10 in the same culture medium and cultured to an optical density (OD) of 1, measured at 600 nm on a Philips PU8740 spectrophotometer (brand name).

10

The expression was then induced either thermally (K38) at 42°C for 15 min, or chemically (BL21(DE3)) by adding 1 mM IPTG. It was continued for 3-5 hours at 37°C. The OD_{600nm} of the cultures was measured at various times.

15

At the end of the expression, a volume of culture containing the equivalent of 0.1 OD of bacteria was removed. The bacteria were harvested by centrifugation and suspended in 50 µl of lysis solution (LS: 50 mM

20   Tris-Cl, pH 8.0, 2.5 mM EDTA, 2% SDS, 4 M urea, 0.7 M β-mercaptoethanol). After a few minutes at ambient temperature, 10 µl were loaded onto a 16.5% polyacrylamide gel for "Tricine" type electrophoresis [23], which makes it possible to obtain good separation

25   of low molar mass proteins.

In Figure 4:
Panels **A, C** and **E**: The bacteria were transformed with the plasmids pT7-7. pT7-7-TME1, pT7-7-TME2 (panel A),

30   pGEXKT, pGEXKT-TME1, pGEXKT-TME2 (panel C), and pGEXKT-dp-TME1 and pGEXKT-dp-TME2 (panel E), and then cultured and induced as described above. The bacterial growth was followed by measuring the increase in turbidity of each culture by measuring the optical density at 600 nm

35   as a function of the time in hours.

Panels **B**, **D**, **F**: The bacteria were sampled at the time indicated in the text and treated as described above. They were then deposited onto an electrophoresis gel,
5 either 16.5% acrylamide of the "Tricine" type (panel B), or 14% acrylamide of the Laemmli SDS-PAGE type (panels D and F). The electrophoresis shown in panel F migrated for a longer period of time than that shown in panel D, in order to improve the separation of the
10 bands in the 30 000 Da region. After migration, the gels were stained for 10 minutes with Coomassie blue in a solution of 40% methanol, 10% acetic acid and 0.1% Coomassie blue R250, and then destained in a solution of 10% methanol, 10% acetic acid and 1% glycerol.
15

Whatever the system tested, the first observation is that the frequency of transformation of the bacteria was low. For the bacteria that could be selected, the result of the expression tests was systematically
20 negative. An example is given in Figure 4, panels A and B, with the series BL21(DE3)pLysS {[pT7-7], [pT7-7-TME1] or [pT7-7-TME2]}. As illustrated by comparing the growth curves of panel A of Figure 4, the inventors noted, with the clones transformed with pT7-7-TME1 or
25 pT7-7-TME2 and resistant on solid medium, that the induction stops the bacterial growth virtually immediately, unlike the clones containing the plasmid alone. Similarly, as can be seen in Figure 4(B), no band of proteins migrating in the region corresponding
30 to the molecular mass of the expression products (~ 3-4000 Da) or of oligomers thereof ({1, 2, 3, etc.}) × molecular mass) can in fact be observed.

The most probable explanation for this situation is
35 that the expression of the membrane domains is very

toxic for the bacterium. The difficulty in obtaining transformants implies that a base line expression, even very low, is sufficient to kill them. It also shows that the pLysS system is not perfect for preventing
5 this base line expression. Among the bacteria that withstand the transformation step, the induction of expression of the hydrophobic domains becomes immediately lethal. The systems used effectively make it possible to protect the host bacterium against a
10 base line expression, but as soon as this expression is induced, the toxicity is immediate and the bacteria are killed.

Example 3: Expression of sequences encoding the GST-
15 ## TME1 and GST-TME2 fusion proteins

The expression vectors were constructed as described in Example 1, and then introduced into the BL21(DE3)pLysS bacteria. The BL21(DE3)pLysS bacteria were used in the
20 interests of comparison with the preceding experiments since the expression of GST or of its chimeras does not require the DE3-pLysS system.

The expression was induced with IPTG as for that of the
25 domains alone. The characteristics of the proteins produced are summarized in Table 2 below.

Table 2

Plasmid	Chimera, abbreviation	Construct	Size, aa	Mass Da
pGEXKT	GST, G	$_1M-D_{239}$	239	27469
pGEXKT-T1	GST-TME1, GT1	$_1M-S_{233-347}M-A_{383}$	269	30506
pGEXKT-T2	GST-TME2, GT2	$_1M-S_{233-717}E-A_{746}$	263	30191

The amino acids (aa) are indicated with the one-letter code. The numbering of the sequences is done with respect to the proteins of origin, GST and viral
5   polyprotein. That which refers to the membrane domains is indicated in italics.

Panels C and D of the attached Figure 4 show the results obtained. The growth curves for the bacteria
10 transformed with the various plasmids show that expression of the GT1 and GT2 chimeras is toxic. As can be seen on the electrophoresis gel of the Laemmli SDS 14% PAGE type [24], the expression of TME1 fused to GST is accompanied by the absence of a band migrating at
15 the expected size of 30 kDa. This implies that a very low level of expression of the chimera is sufficient to kill the bacteria. On the other hand, the GST-TME2 chimera is this time visible on the electrophoresis gel, in the region of expected molecular mass of
20 30 kDa. The level of expression remains limited however.

The protein produced is not soluble despite the presence of GST in the fusion. In fact, as shown in the
25 attached Figure 6, the solubilization, folding and purification trials for the GST-TME2 chimera were a failure.

To obtain the results represented in this Figure 6, the
30 GST and GST-TME2 proteins were expressed as described in Figure 4, using 150 ml of culture medium. The bacteria were then harvested by centrifugation and suspended (20 mM KPO_4, pH 7.7, 0.1 M NaCl, 1 mM EDTA, 1 mM NaN_3) so as to have 100 OD/ml. Two ml of each
35 culture were removed for sonication with 30 sec pulses

at an amplitude of 15%. After sonication, a sample is taken for electrophoresis. It corresponds to the well "To" in Figure 6 (corresponding to the "total").

5   A first low-speed centrifugation (5000×g, 15 minutes) makes it possible to separate the non-ruptured bacteria and the inclusion bodies from the soluble or membrane proteins. The latter are found in the supernatant and a sample is taken. It corresponds to the well "Surn" in
10 Figure 6.

The fraction containing GST alone is then treated with an affinity resin that makes it possible to bind and then elute specifically this protein (well "Af" of the
15 GST gel in Figure 6).

The fraction containing the non-soluble GST-TME2 protein is treated either with a mild detergent such as triton X100 (TX100), in the presence or absence of
20 NaCl, or with a more solubilizing but more destructuring detergent such as sarkosyl, before again being diluted in TX100 and passed over affinity resin.

The results in Figure 6 show that GST is present in the
25 soluble fraction, unlike the GST-TME2 fusion, which indicates that the latter is insoluble. The supernatant containing the GST is passed over an agarose-GSH resin capable of binding GST. This GST is then eluted with an excess of GSH (well marked "Af" of the GST gel in
30 Figure 6).

The pellet containing the GST-TME2 fusion is not solubilized in the presence of a mild detergent such as TX100 (with or without added NaCl, well "TX100 +/-
35 NaCl" of the GST-TME2 gel), but it can be solubilized

with a more aggressive detergent such as sarkosyl.
However, after dilution of the protein thus solubilized
in TX100, a mild detergent which should favour its
folding, the protein is not retained on the affinity
5   resin, unlike GST, which suggests that the fusion
protein cannot be folded.

These tests clearly indicate that the GST-TME2 protein
is produced in the form of inclusion bodies that cannot
10   be correctly folded.

**Example 4: Expression of expression vectors encoding
the fusion proteins including an Asp-Pro site and a GST
site**

15

The construction of the vectors was carried out as
described above and for the two vectors encoding the
GST-TME1 and GST-TME2 chimeric proteins, so as to
produce the vectors encoding the GST-Asp-Pro-TME1 and
20   GST-Asp-Pro-TME2 chimeric proteins. They are summarized
in Table 3 below.

Table 3

Plasmid	Chimera, abbreviation Fig. 4	Construct	Size, aa	Mass, Da
pGEXKT-dp-$T1$	GST-DP-$TME1$; $G_{DP}T1$	$_1$M-D$_{233}$-dp-$_{347}$M-A$_{383}$	271	30718
pGEXKT-dp-$T2$	GST-DP-$TME2$; $G_{DP}T2$	$_1$M-S$_{233}$-dp-$_{717}$E-A$_{746}$	265	30403

25   The amino acids (aa) are indicated with the one-letter
code. The numbering of the sequences is done with
respect to the proteins of origin, GST and viral
polyprotein. That which refers to the membrane domains

is indicated in italics.

The vectors were tested as described in the preceding paragraph. The results obtained are shown on panels E and F of the attached Figure 4.

The growth curves for the bacteria transformed with the various plasmids show that the expression of the $G_{dp}T1$ and $G_{dp}T2$ chimeras is clearly less toxic than in the previous cases. Panel F shows that, this time, TME1 is produced due to the presence of the DP cleavage site. Its level of expression, as can be seen in panel F, is relatively moderate, but significant. GST-DP-TME2 is clearly overproduced. The two proteins migrate in their expected molecular mass region.

The effect of the addition of the DP dipeptide is as significant as it is unexpected: it amplifies the expression of the domains and suppresses their toxicity. This effect of attenuation of the toxicity is not known for the DP dipeptide, the only property of which that has been reported to date is its ability to be cleaved by formic acid. Since the effect is observed on two different peptides that are both initially toxic for the bacterium, it is therefore reasonable to think that this property may extend to other hydrophobic and toxic peptides.

The inventors verified that the site can be effectively cleaved by formic acid: the cleavage is slow and requires approximately 7 days at ambient temperature.

The assays of expression at low temperature (20°C) overnight of these chimeras made it possible to demonstrate that they are produced in native form. In

fact, it is possible to detect GST transferase activity in the membrane fraction of the bacteria. In addition, this activity is measured in solution when the membranes are solubilized in the presence of a non-
5   ionic detergent such as β-D-dodecylmaltoside, after centrifugation.

Example 5: Expression of expression vectors encoding the fusion proteins including an Asp-Pro site and a
10   site encoding thioredoxin (TrX)

The pET32a-TrX, pET32a-TrX-dp-TME1 and pET32a-TrX-dp-TME2 expression vectors were constructed as described above and were then introduced into BL21(DE3)pLysS
15   bacteria. The BL21(DE3)pLysS bacteria were used in the interests of comparison with the previous experiments since the expression of GST or of its chimeras does not require the DE3-pLysS system. The positive clones were cultured and induced as described above.
20

The induction of expression was carried out with IPTG, as for that of the domains alone. The characteristics of the proteins produced are summarized in Table 4 below.
25

Table 4*

Plasmid	Chimera, abbreviation Fig. 4	Construct	Size, aa	Mass, Da
pET32a	Thioredoxin; TrX	$_1M-C_{189}$	189	20397
pET32a-Gend-dp-T1	TrX-DP-TME1; $T_{DP}T1$	$_1M-S_{115}-PK-$Gend-dp-T_1	171	17796
pET32a-Gend-dp-T2	TrX-DP-TME2; $T_{DP}T2$	$_1M-S_{115}-PK-$Gend-dp-T_2	165	17481

* : T1 = TME1 and T2 = TME2

The amino acids (aa) are indicated with the one-letter code. The numbering of the sequences is done with respect to the proteins of origin, GST and viral polyprotein. That which refers to the membrane domains is indicated in italics. "Gend" refers to the C-terminal sequence of the GST originating from the constructs with the pGEXKT plasmid. It corresponds to the primary peptide sequence SDLSGGGGGLVPRGS. The thioredoxin-SDLSGGGGGLVPRGS-DP-(TME1 or TME2) chimeras are shorter than the protein encoded in the vector of origin since the insertion is effected immediately after the thioredoxin.

In Figure 5:

A: the bacterial growth was followed by measuring the increase in turbidity of each culture by optical density at 600 nm as a function of time.

B: the bacteria were sampled as indicated for Figure 4. They were then loaded onto a Laemmli SDS-PAGE type 14% acrylamide electrophoresis gel and treated as indicated for Figure 4.

As expected, and as shown by the growth curves represented in the attached Figure 5A for the bacteria transformed with the various plasmids, expression of the TrX-DP-TME1 and TrX-DP-TME2 chimeras according to the present invention is not toxic. The Laemmli 14% SDS-PAGE [24] electrophoresis gel represented in the attached Figure 5B shows that each chimera is overproduced.

The present invention therefore makes it possible to produce, by genetic recombination, hydrophobic peptides

corresponding to the membrane domains of the E1 and E2 proteins of the hepatitis C virus envelope, the expression of which was acknowledged to be lethal in the techniques of the prior art. In addition, since the

5   effect is observed on two peptides that are really different and both initially toxic for the bacterium, this indicates that the present invention concerns other hydrophobic and toxic peptides.

10   **Example 6: Effect of the DP dipeptide on the toxicity of the TME1 and TME2 transmembrane domains expressed without fusion protein in the bacterium**

     This example makes it possible to evaluate the antitoxic

15   effect of the DP dipeptide inserted in the absence of GST or TrX fusion protein in accordance with the attached Claim 1.

     A) Materials: The pT7-7-pt_{TME1} and pT7-7-pt_{TME2} plasmids

20   are those which are described in Example 1. The pT7-7-dp-pt_{TME1} and pT7-7-dp-pt_{TME2} plasmids were constructed and cloned in pT7-7 (SEQ ID No. 33) as described in Example 1, but using the Nde I (5′) EcoR I (3′) sites of the plasmid. The upstream (5′) oligonucleotides

25   integrate the dp sequence (gacccg) after the 1st methionine (atg). The matrices used to generate each DNA were the pT7-7-pt_{TME1} and pT7-7-pt_{TME2} plasmids. The sequences were verified after cloning.

30   The oligonucleotides are as follows:

35

i) Cloning of the sequence encoding (M)DP-TME1 in pT7-7:

OL19 (+) : 5'-CG<u>CATATG</u>GACCCGATCGCTGGTGCT - 3' (*Nde* I underlined) = (SEQ ID No. 15 of the attached sequence listing);

OL20 (-) : 5'-<u>GAATTCC</u>TAAGCGTCAACACCAGC-3' (*EcoR* I underlined) = (SEQ ID No. 16 of the attached sequence listing).

ii) Cloning of the sequence encoding (M)DP-TME2 in pT7-7:

OL28 (+) : 5'-CG<u>CATATG</u>GACCCGGAATACGTTGTTC-3' (*Nde* I underlined) = (SEQ ID No. 26 of the attached sequence listing);

OL29 (-) : 5'-CAG<u>AATTCC</u>TAAGCTTCAGCCTGAGAG-3' (*EcoR* I underlined) = SEQ ID No. 27 of the attached sequence listing).

The pT7-7-*dp-pt*$_{TME1}$ and pT7-7-*dp-pt*$_{TME2}$ expression vectors obtained are given in the attached sequence listing (SEQ ID No. 44 and SEQ ID No. 45).

B) Legend of the attached figures 7A and B: the bacterial strain BL21(DE3)pLysS was transformed either with the plasmid alone or with the various versions of pT7-7 integrating the 4 constructs expressing TME1, M-DP-TME1 (Figure 7A), or TME2, M-DP-TME2 (Figure 7B). M represents methionine; it is present at the N-terminal position of the peptides when the toxic proteins are produced according to the present invention with the pT7-7 plasmid.

The growth of the various clones was compared after induction with IPTG, according to the protocol identical to the chemical induction described in Example 2, and averaged over the OD values of 4 different clones for
5   each construct.

C) Results:

Figures 7A and 7B show that the bacteria that have a
10   plasmid expressing TME1 and TME2 proteins grow less rapidly after induction than the control strain which is transformed with the pT7-7 vector alone.

These results show that the strains transformed with the
15   plasmids expressing the M-DP-TME1 (SEQ ID No. 50) and M-DP-TME2 (SEQ ID No. 51) versions according to the invention grow significantly better than those that express the TMs without DP. This is true for TME1, and even more clearly so for TME2.
20
The conclusion is that the N-terminal insertion of DP in accordance with the present invention contributes, surprisingly, to a significant decrease in toxicity of the expression of the membrane domains, in particular in
25   the absence of a soluble fusion protein such as GST or thioredoxin.

Reference list

[1] Christendat D., Yee A., Dharamsi A., Kluger Y., Gerstein M., Arrowsmith C.H., and Edwards A.M., (2000), *Prog. Biophys. Mol. Biol.* **73**, 339-345 ;

[2] Hammarstrom M., Hellgren N., Van Den Berg S., Berglund H., and Hard T., (2002), *Protein Sci.* **11**, 313-321 ;

[3] Falson P. (1992), *Biotechniques* **13**, 20-22 ;

[4] Falson P., Penin F., Divita G., Lavergne J.P., Di Pietro A., Goody R.S., and Gautheron D.C. (1993), *Biochemistry* **32**, 10387-10397 ;

[5] Ciccaglione A.R., Marcantonio C., Costantino A., Equestre M., Geraci A. and Rapicetta M. (2000) *Virus Genes* **21**, 223-226 ;

[6] Sisk W.P., Bradley J.D., Kingsley D., and Patterson T.A. (1992) *Gene* **112**, 157-162 ;

[7] Paulsen I.T., Sliwinski M.K., Nelissen B., Goffeau A., and Saier M.H. Jr. (1998) *FEBS Lett* **430**, 116-125 ;

[8] Decottignies A. and Goffeau A. (1997) *Nat Genet* **15**, 137-145 ;

[9] Arechaga I., Miroux B., Karrasch S., Huijbregts R., de Kruijff B., Runswick M.J. and Walker J.E. (2000) *FEBS Lett* **482**, 215-219 ;

[10] Miroux B. and Walker J.E. (1996) *J. Mol. Biol.* **260**, 289-298 ;

[11] Mayo M.A., and Pringle C.R. (1998) *J. Gen Virol.* **79** (Pt4), 649-657 ;

[12] Op De Beeck A., Montserret R., Duvet S., Cocquerel L., Cacan R., Barberot B., Le Maire M., Penin F. and Dubuisson J. (2000) *J Biol Chem* **275**, 31428-31437 ;

[13] Choo Q.L., Kuo G., Weiner A.J., Overby L.R., Bradley D.W., and Houghton M. (1989) *Science* **244, 359-362** ;

[14] Ciccaglione A.R., Marcantonio C., Costantino A., Equestre M., Geraci A. and Rapicetta M. (1998) *Virology* **250**, 1-8 ;

[15] Ciccaglione A.R., Marcantonio C., Equestre M., Jones I.M. and Rapicetta M. (1998) *Virus Res* **55**, 157-165 ;

[16] Op De Beeck A., Cocquerel L., and Dubuisson J. (2001) *J Gen Virol* **82**, 2589-2595 ;

[17] Sharp P.M., Cowe E., Higgins D.G., Shields D.C., Wolfe K.H., and Wright F. (1998) *Nucleic Acids Res* **16**, 8207-8211 ;

[18] Mullis K.B., and Faloona F.A. (1987) *Methods Enzymol* **155**, 335-350 ;

[19] Tabor S. and Richardson C.C. (1985) *Proc Natl Acad Sci* USA **82**, 1074-1078 ;

[20] Guan K.L., and Dixon J.E. (1991) *Anal Biochem* **192**, 262-267 ;

[21] Hakes D.J., and Dixon J.E. (1992) *Anal Biochem* **202**, 293-298 ;

[22] Tabor S. (1990) in *Current Protocols in Molecular Biology*, pp. 16.12.11-16.12.11, Greene Publishing and Wiley-Interscience, New York ;

[23] Schagger H. and von Jagow G. (1987) *Anal Biochem* **166**, 368-379 ;

[24] Laemmli U.K. (1970) *Nature* **227**, 680-685.

[25] Sambrook, Fritsch and Maniatis, Molecular cloning, A laboratory manual, second edition, Cold spring Harbor Laboratory Press, 1989.